Physical Abilities and Military Task Performance: A Replication and Extension

Ross R. Vickers, Jr.

Naval Health Research Center

Document No. 09-30

This research has been conducted in compliance with all applicable federal regulations governing the protection of human subjects in research.

Naval Health Research Center
140 Sylvester Rd.
San Diego, California 92106-3521

Physical Abilities and Task Performance: Predicting Military Task Performance

Ross R. Vickers, Jr.
Naval Health Research Center
140 Sylvester Road
San Diego, CA 92106-3521

This represents report 09-30 and was supported by the U.S. Department of Defense under work unit 60704. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government. Approved for public release; distribution is unlimited. This research has been conducted in compliance with all applicable federal regulations governing the protection of human subjects in research.

Abstract

Strength influences the performance of military physical tasks. These influences can be summarized by models that treat strength as a general dimension that affects performance on tasks in general. Previous findings also indicate that combining a general strength (GS) dimension with an aerobic capacity (AC) dimension yields a model that accounts for the full pattern of association between physical ability tests and lifting and carrying. This study attempted to replicate the earlier findings using a strength test battery with some new strength measures, a different set of military tasks, and a different military population. Structural equation models were constructed to represent strength as a single construct, a two-dimensional construct based on measurement modality, and a seven-dimensional model based on specific functional movements. Performance was represented as a single general performance dimension that added digging and casualty evacuation to the manual materials-handling tasks that had been previously studied. A modified unidimensional model maximized the prediction of performance. Adding AC to the strength model improved performance prediction, but adding muscle endurance (ME) and anaerobic power (AP) did not. The results provided a very close replication of earlier findings while extending the model to a wider range of military tasks and a new population.

Strength influences the performance of many physical tasks. The strength–performance association can be modeled by representing strength as a single general factor and performance as a general factor (Vickers, 1995, 1996, 2003a; Vickers, Hodgdon, & Beckett, 2009). Such models have produced strength–performance correlations ranging from r = .32 to r = .96. Typical values fall in the middle of this range, but the model has explained the overall pattern of association of specific strength tests with the performance of specific tasks in every instance. The strong correlations between the general factors combined with the lack of substantial residual associations indicate that a general strength (GS) dimension adequately summarizes the results of strength testing when predicting performance. Specific strength measures do not have to be matched to specific tasks.

The apparent importance of general muscular strength for task performance may be misleading. Two studies have shown that the association of GS with task performance is weaker when other physical abilities are considered (Vickers, 2003a; Vickers et al., 2009). The association of strength with performance may be inflated because some of the effects of other causal variables are attributed to GS. This inflation would be an example of omitted variable bias (James, Mulaik, & Brett, 1982). The most recent effort indicated that the combination of GS and aerobic capacity (AC) was sufficient to represent a range of more complex models that could have been generated by adding muscle endurance (ME) and/or anaerobic power (AP) to the model. If this finding can be replicated, the potential for omitted variable bias would be substantially reduced since ME and AP could be eliminated as potential sources of bias.

The tasks examined also limit the inferences that can be drawn from previous studies. Nearly all of the tasks have been manual materials-handling tasks involving lifting, carrying, pushing, pulling, and so forth. While such tasks capture much of the physical variation in physical demands of military jobs (e.g., Beckett & Hodgdon, 1987; Rayson, Holliman, & Belyavin, 2000; Robertson & Trent, 1982, 1985, Singh et al., 1991), they may not represent the full range of tasks that are important in the military. Critical task analyses often lead to the conclusion that other tasks such as digging and load carriage are essential for military effectiveness (Rayson et al., 2000; Singh et al., 1991).

This report extends the previous modeling of physical abilities and performance in four directions. First, the effects of the strength measurement modality are examined. One previous study of modality effects produced ambiguous results because isokinetic, isometric, and isoinertial strength measures all predicted task performance and none were clearly superior to the others (Dempsey, Ayoub, & Westfall, 1998). Second, a model based on GS is compared with a model with a number of more specific strength dimensions (e.g., handgrip). This topic is explored because it is reasonable to think that accurate task prediction might require matching tests to tasks. Third, the problem of omitted variable bias is explored further with new measures of ME. Fourth, coverage of the task domain has been increased. Most of the earlier work has been done with tasks that require a brief maximal effort (Vickers, 1995, 1996). Subsequent models have introduced longer-lasting tasks, but these tasks have been mixed with short maximal performance tasks (Vickers, 2003a; Vickers et al., 2009). The task set in this study emphasizes efforts that extend over several minutes. The study goal was to replicate the previous findings, if replication was possible, with these several extensions that substantially broadened the scope of the area of inquiry.

Methods

Sample

The sample consisted of 116 male soldiers. Their average age was 25.5 years (*SD* = 5.8). Their average weight was 79.4 kg (*SD* = 10.6). Their average height was 177.4 mm (*SD* = 7.6). (see Singh et al., 991, p. 242). Only 88 of these participants completed the field test battery, so the sample size was set at 88 for the present structural equation model (SEM) analyses.

Physical Ability Tests

Aerobic Capacity. The test protocol for assessing AC required subjects to walk on a treadmill at 88.9 m/min while carrying a 24.5-kg fighting order load. The test began with a 2-min warm-up at 0% incline. Following the warm-up, the incline of the treadmill was increased 2% every 3 min until anaerobic threshold was reached. The incline then was increased 2% every minute until maximal oxygen uptake was reached. No specific criteria for determining that a valid maximum had been achieved were reported, and there was no indication of whether the measurement relied on a plateau in oxygen uptake or was a peak value.

Anaerobic Power. AP measurements consisted of 30 s of "supramaximal" bouts on modified Monark ergometers. The modifications provided an interface with a computer system that calculated and provided resistance loads based on the subject's body weight. The exercise bout consisted of a warm-up period and the test period. Three seconds after warm-up, subjects were instructed to increase their pedaling speed to maximum. Load was applied at maximum and subject worked for 30 s with encouragement given in the last 5 s. A computer recorded the test resistance, revolutions per minute (RPM), peak power output, mean power output, power decline (i.e., the percentage decrease from the peak power output to the end of the bout), and total work during the bout. Separate AP tests were performed for the legs and the arms.

Isometric Strength Tests. Strength measurements were made with a system that required the test subject to exert force on a bar or handle that was attached by a cable to a load cell.

> *Handgrip*. A handgrip dynamometer was used to determine the maximum grip strength of each hand.
>
> *Arm flexion*. The subject grasped the bar at shoulder width with the arm at a 105° angle.
>
> *Trunk flexion*. A shoulder harness connected to a load cell via two pulleys was used. The test was executed at a hip angle of 160°. The chain that connected the apparatus to the load cell was at the test subject's back.
>
> *Trunk extension*. Participants stood on a platform with the lateral borders of the feet at shoulder width. The individual assumed a lifting position holding the bar with an over-and-under grip. The subject's arms were fully extended, and his hips were flexed at a 160° angle. Maintaining that flexion, the subject then pulled up using the muscles in his back while keeping his back straight.

Isokinetic-Concentric Strength Tests. For isokinetic tests, a motor was attached to the cable to control the rate of movement of the bar. A goniometer was used to control the angles for the range of motion.

Arm flexion. Subject grasped the bar at shoulder width with arms straight (180°) at the beginning of the test. Subject performed a concentric contraction bringing the arms to a 40° angle with an angular velocity of 30°/s. Thus, a contraction lasted 4.67 s.

Leg extension. Subject began in a standing position with feet shoulder-width apart. He then flexed his legs to a 90° angle and grasped a handle that was adjusted to his height. The subject then stood up, extending the knees to a 180° angle, and generating as much force as possible during the process. The movement was 30°/s, so exertion lasted 3 s.

Trapezius lift. Subject stood with feet at shoulder width grasping handles that were 38.5 cm apart to mimic the grip used in lifting ammunition boxes. At the start of the motion, the arms were fully extended. Subject then raised his arms to clavicle level at a rate of 30°/s. The duration of the contraction varied from subject to subject because of height differences.

Bench press. Subject was supine on a bench. The bar connected to the load cell was set at 1 inch above his sternum. Subject extended his arms to their full extension at a rate of 30°/s.

Trunk extension. Subject began with body flexed at the hips to 150°. Keeping his back and legs straight, the subject then straightened up to 170°. The test was performed at an angular velocity of 15°/s, so extension took 1.3 s.

Trunk flexion. Subject began with body slightly flexed at the hips to form an angle of 170°. Subject then bent forward keeping back and legs straight until the angle diminished to 150°. The test was performance at an angular velocity of 15°/s, so extension took 1.3 s.

Knee extension. The knee extension test was performed on a Cybex dynamometer (Cybex International, Inc., Medway, MA). The subject was seated with his knee at a 90° angle. He then straightened his leg to a 180° angle. The movement was at an angular velocity of 180°/s. The knee extension test was performed separately for the right and left legs.

Knee flexion. The knee flexion test was performed on a Cybex dynamometer. The subject was seated with his knee at a 180° angle. He then flexed his leg to a 90° angle. The movement was at an angular velocity of 180°/s. The knee flexion test was performed separately for the right and left legs.

Muscle Endurance. The four ME tests assessed the ability to sustain submaximal muscular exertion. For three of the tests, the exertion was continuous. Performance was the length of time that the required exertion could be sustained. For the fourth test, the exertion was

intermittent and was repeated until the test subject was unable to continue. The specific tests were:

Handgrip. Using the handgrip dynamometer, subjects were directed to generate a grip of 21 kg and to maintain it as long as possible. The target force was chosen to equal the weight of a full jerry can. Subjects were given encouragement during the test. The test was stopped when the force of the muscle contraction dropped below the targeted value and the participant was no longer able to return to the required value within 2 s. Separate tests were performed for the right and left hands.

Isometric arm flexion. Subjects held a free-weight bar weighing 20 kg with their arms at a 105° angle as determined by a goniometer. The test stopped when the arm angle could no longer be maintained even with encouragement.

Trapezius lift. The subject stood erect with his feet shoulder-width apart. He held a 20.9-kg load with his arms at his sides and fully extended. He then lifted the 20.9-kg load to clavicle height at a rate of 6 times per minute. The lifting rate was controlled by a metronome. The lifts continued until the subject was unable to maintain the required pace or had performed 100 lifts.

Field Performance Test Battery

The field performance test battery was constructed to represent a range of critical tasks that might be performed in combat (Singh et al., 1991). Experienced military personnel identified the critical tasks. The specific tasks employed by Singh et al. (1991) were:

Digging Slit Trenches. A metal box with dimensions 1.8 m long x 0.6 m wide x 0.45 m deep was filled with standardized gravel <1 cm diameter to a total volume of 0.5 m^3. Participants were instructed to shovel the gravel at the maximum rate possible. The task ended when all of the gravel had been removed. The instructions to participants suggest that this included some sweeping up of the last remains by hand until it was no longer possible to pick up a handful of gravel. The test score was the time in seconds.

Casualty Evacuation. Subject was required to evacuate another individual of approximately the same height and weight over a distance of 100 m. The fireman’s carry was used to transport the casualty. During the test, the test subject wore a uniform and carried a weapon and webbing. This test was performed with maximal effort. The test score was the time in seconds.

Jerry Can Carry. Subject picked up a 21-kg jerry can of water, carried it 35 m, emptied it into a funnel at the height of a truck bed (1.3 m), and ran back to the starting line. He then picked up another can and repeated the process. A trial consisted of 3 roundtrips to and from the starting line. The time for the event was recorded when the foot touched the starting line after the third carry. The test score was the time in seconds.

Handle Manual Material. Ammunition boxes weighing 20.9 kg were lifted from the floor and placed on a flat surface 1.3 m above the floor. The distance lifted equaled the height of a lift to a truck bed. The test consisted of performing 48 such lifts. Subjects wore a Polar Sport Tester heart rate monitor (Polar Electro Inc., Lake Success, NY) and were instructed to perform the series of lifts at a rate that equaled 70% of maximal aerobic power. A submaximal work rate was adopted to minimize the risk of injury. The test score was the time in seconds.

The original study included a weight-loaded march as a fifth task. Soldiers marched at a pace of 88.9 m/min while carrying a 24.5-kg load. Subjects continued at the set speed for 16 km or until they could no longer maintain the required pace. The test score was the distance covered. Because 77.3% of all participants completed the full 16 km, the variation in performance was too restricted to analyze the relationships to ability measures with confidence. This performance measure was dropped from the analysis.

Analysis Procedures

The correlation matrix and descriptive statistics for all the measurements was extracted from Singh et al. (1991; most of the correlation matrix can be found on pp. 278–280 of that report). This information was used to construct the covariance matrix that was analyzed. Statistics were reported to two decimal places, so it is possible that limited precision had an effect on the model evaluations presented later in this report. Appendix A provides the relevant descriptive statistics from Singh et al. (1991).

The LISREL 8.5 program (Scientific Software International, Chicago, IL) developed by Joreskog and Sorbom (1996) was employed to estimate structural models. Anderson and Gerbing's (1988) two-step approach to modeling was adopted. The construction of measurement models for ability and performance was the first step. The scaling of the latent traits in these models was established by fixing the latent trait loading at 1.000 for one test or task on each hypothesized dimension.

The estimation of relationships of physical abilities with task performance was the second analysis step. These analyses were carried out with the parameters of the measurement models fixed at the values determined when developing the measurement models. Conceptually, this two-step procedure reduces the ambiguity of research findings by ensuring that negative results are not merely manifestations of poor measurement models (Meehl, 1990). Also, this approach reduces the risk that a good measurement model will mask poor fit in the substantive model (McDonald & Ho, 2002).

Models were evaluated by χ^2 tests that compared the χ^2 for the model with the χ^2 for the null model. The root mean square error of approximation (RMSEA; Browne & Cudeck, 1993; Steiger, 1990) and the nonnormed fit index (NNFI, Bentler & Bonett, 1980; Tucker & Lewis, 1973) were additional model evaluation criteria (cf., Arbuckle & Wothke, 1999, Appendix C, for details).

Table 1. Performance of Alternative Strength Measurement Models

Logical	Min *t*	Max *t*	*df*	χ^2	RMSEA	NNFI
Unidimensional	4.81	8.38	90	358.86	.19	.52
Modality	4.91	8.23	89	354.49	.19	.52
Seven[a]	5.06	12.83	69	117.04	.09	.81
Seven [a]	5.39	12.81	69	104.42	.08	.82
Empirical						
Two	5.49	10.84	89	248.76	.15	.62
Three	5.42	10.91	87	194.41	.12	.72

Note. The minimum and maximum *t* values for individual factor loadings indicate that all of the strength tests met the accepted criterion for justifying that they were acceptable indicators of the latent trait to which they were assigned.
[a]Leg extension was assigned to the bench press/lat pull-down dimension in this model.
[b]Leg extension was assigned to the trunk flexion/trunk extension dimension in this model.

Exploratory factor analyses were conducted to develop alternatives to the conceptual strength measurement models. These analyses were conducted with SPSS-PC, version 16 (SPSS, Inc., Chicago, IL). Kaiser's (1960) criterion was one basis for deciding how many factors to extract. O'Connor's (2000) implementation of Horn's (1965) parallel analysis criterion was a second basis for deciding how many factors to extract. The pattern loadings from an oblimin rotation were used to assign tests to factors (cf., Gorsuch, 1983). An oblique rotation was chosen because previous research suggested that strength factors tend to correlate (Vickers, 2003b). If this were not the case in the present data, the oblimin rotation would produce factors with very low correlations. The oblique rotation provided the opportunity to identify a model with orthogonal dimensions, if appropriate, without assuming that the dimensions should be orthogonal.

Results

Measurement Modality Model. Measurement modality had little effect on strength measurement (Table 1). This point was established by comparing a unidimensional strength model to a two-dimensional model that was labeled the measurement modality model. One latent trait in the modality model was defined by the set of static isometric strength measures. The other latent trait was defined by the dynamic isokinetic measures.

Two findings suggested that the distinction between static and dynamic strength was likely to be unimportant. First, the two latent traits were very highly correlated ($r = .93$). Second, the modality model did produce a significant improvement in the accounting for the observed strength test covariations, but the improvement was modest in absolute magnitude ($\Delta\chi^2 = 4.37$, 1 *df*, $p < .037$).

Specific Strength Model. A specific strength model was constructed to assess the claim that narrow strength factors will predict task performance better than broad strength dimensions. The initial specific strength model contained seven latent traits that were defined a priori:

dynamic and static trunk extension; dynamic and static arm flexion; bench press, trapezius lift, leg extension; dynamic and static trunk flexion; right and left handgrip; right and left knee

Table 2. Strength Measurement Models from Exploratory Factor Analysis

	Number of Factors in Solution					
	1	2		3		
	A	A	B	A	B	C
Trunk extension (D)	.817	.799		.738		
Left knee extension (D)	.791		-.763		-.734	
Left knee flexion (D)	.764		-.877		-.869	
Right knee extension (D)	.763		-.759		-.724	
Trunk extension (S)	.748	.658		.610		
Right knee flexion (D)	.706		-.925		-.895	
Trunk flexion (D)	.674	.511				-.726
Trapezius lift (D)	.654	.691		.705		
Trunk flexion (S)	.647	.491				-.879
Arm flexion (S)	.645	.699		.666		
Left handgrip (S)	.617	.398		.491		
Right handgrip (S)	.603		-.373	.419	-.395	
Bench press (D)	.593	.726		.708		
Leg extension (D)	.558	.491		.455		
Arm flexion (D)	.541	.725		.700		

Note. Table entries are the pattern loadings from an oblimin factor solution (cf., Gorsuch, 1983). Loadings <.30 (absolute) have been dropped to facilitate identification of the factor structure. "D" indicates a dynamic strength measure; "S" indicates a static strength measure.

flexion; and right and left knee extension. Note that with one exception, the dimensions in the model were defined either by tests that involved dynamic and static measures of a single muscle action or the same action performed by the right and left muscle groups.

The initial model produced a substantial improvement in goodness of fit relative to the unidimensional model ($\Delta\chi^2$ = 237.45, 20 *df*, $p < .001$). The modification indices from the initial analysis of the a priori model indicated that goodness of fit would improve if leg extension were aligned with the trunk flexion dimension instead of its original alignment with the bench press and trapezius lift. This shift improved on the overall fit by 5.3% ($\Delta\chi^2$ = 12.62). The original alignment of leg extension was speculative, so leg extension was reassigned in the final specific strength model.

Exploratory Factor Analyses

Exploratory factor analysis produced three principal components with $\lambda_i > 1.00$ (λ_1 = 7.44, λ_2 = 1.43, λ_3 = 1.22). The first principal component accounted for 49.6% of the variance; the first three principal components accounted for 67.3% of the variance. Table 2 presents the factor structure for models with one, two, and three dimensions.

The SPSS routine developed by O'Connor (2000) provided a parallel analysis criterion (cf., Horn, 1965) for the number of factors. The 95th percentiles for the first, second, and third components were 1.82, 1.61, and 1.48, respectively. The 50th percentiles were 1.67, 1.51, and 1.39, respectively. By reasonable standards, the second and third factors would be discounted as chance findings. However, for the present purposes, the $\lambda_i > 1.00$ rule was applied to explore the

Table 3. Latent Trait Correlations: Seven-Dimensional Model

Bench press/ trapezius lift	1.00						
Trunk extension/ leg extension	0.63	1.00					
Arm flexion	1.04[a]	0.69	1.00				
Trunk flexion	0.56	0.47	0.61	1.00			
Handgrip	0.58	0.42	0.60	0.38	1.00		
Knee flexion	0.59	0.46	0.53	0.48	0.55	1.00	
Knee extension	0.59	0.55	0.65	0.60	0.57	0.83	1.00

[a]Estimated correlations sometimes exceed 1.00 in structural models. This result is attributed to the sampling variability associated with a true correlation of $r \approx 1.00$.

maximum number of plausible factor structures. Accordingly, solutions with 1 to 3 factors were examined.

The relationships among the exploratory factor solutions were simple (Table 2). All strength tests had substantial (>.54) loadings in the unidimensional solution. The two-dimensional solution primarily separated knee strength tests from the general dimension. The knee strength (KS) dimension was strongly related (r = -.674) to the GS dimension defined by the 11 remaining strength tests.

The three-dimensional structure further subdivided the original GS factor. A trunk strength (TS) factor defined by the two trunk flexion tests was added to the two-dimensional model. The KS factor was unchanged from the two-dimensional model. The residual "general" factor was strongly correlated with knee strength (r = -.614) and moderately correlated with trunk flexion strength (r = -.419). KS and TS were moderately correlated (r = .326).

When converted to a structural equation model, the two-dimensional model fit the data better than both the unidimensional and measurement modality models (Table 1). The three-dimensional fit to the data better than any of the simpler models, but the specific strength dimensions model still remained the best option by most criteria.

Table 3 gives the latent trait correlations for the seven-dimensional model. One point to note is that most of the correlations fell in a rather narrow range (r = .38 to r = .65). The exceptions were the correlation of the bench press dimension with arm flexion (r = 1.04) and the correlation of the two knee extension measures (r = .83). The correlation that exceeds 1.00

presumably represents a case in which sampling variation makes a correlation that approaches $r = 1.00$ apparently exceeds the upper limit for correlations.

Other Physical Abilities

The set of ability tests included measures chosen to assess three additional ability constructs, ME, AP, and AC. A measurement model constructed to measure these three dimensions included the 4 ME measures, the 2 peak power measures, and VO_{2max}. The latent

Table 4. Comparison of Alternative Models for Other Abilities

Model	*df*	χ^2	RMSEA	NNFI
VO_{2L}	12	39.43	.17	.79
VO_{2max}	12	29.15	.13	.83

trait scaling was established by assigning a loading of 1.000 to static right-hand endurance, peak power for the leg ergometer, and the AC indicator. The model included only a single AC indicator even though two aerobic measures were available. Maximal oxygen uptake (VO_2) in liters/min (VO_{2L}) and VO_2 in $ml{\cdot}kg^{-1}{\cdot}min^{-1}$ (VO_{2max}) were based on a single measurement procedure. This commonality made it appear wiser to just one of the indicators. In each case, the error variance for this measure was fixed at .000, so the aerobic latent trait corresponded directly to the measured variable.

The choice of an AC indicator affected the fit of the measurement model. The measurement model fit the data better with VO_{2max} as the AC indicator than with VO_{2L}. AC was significantly related to ME ($r = .44$) and AP ($r = .83$) in the VO_{2L} model. The VO_{2max} model produced statistically nonsignificant correlations to ME ($r = .10$) and AP ($r = .16$). Despite the apparent advantage of the VO_{2max} model, both models were employed in parallel when predicting task performance. The parallel analyses made it possible to evaluate total aerobic energy expenditure, VO_{2L}, as a predictor of tasks that required absolute levels of energy expenditure rather without adjusting for size.

Performance Measurement

A unidimensional performance model was constructed initially. The factor loading for casualty evacuation was fixed at 1.000 to set the scale for the latent variable. The residual χ^2 that was less than the degrees of freedom, so RMSEA and NNFI indicated perfect fit. The latent trait loading for digging ($\lambda = 2.56$, $t = 1.67$) did not meet the standard for inclusion as indicators of a general factor. The *t* value for the variance of the latent trait, $t = 1.82$, also fell short of the recommended $t \geq 2.00$.

The unidimensional measurement model was retained even though it did not meet accepted modeling guidelines. If there really were no true latent trait variability, all correlations of the performance latent trait to ability latent traits should equal zero. The variation in slit trench digging performance can be divided conceptually into two parts. The first part is variation arising

from the general ability to perform all of the tasks in the study. The second part is variation that is specific to slit trench digging. If this specific task is related to ability tests or traits independently of the associations that would be predicted from the relationship of general performance to physical ability, those associations would be indicated by the residuals or modification indices for the ability-performance model. Any substantial task-specific associations then could be added to the model.

Table 5. Strength as a Predictor of Performance

Model	Model χ^2	*df*	Change in χ^2	NNFI[a]	Correlation[b]
Unidimensional					
Null	70.61	60	.00	.00	
General strength	55.66	59	14.95	1.00[a]	-.50
Measurement mode					
Null	72.96	60	.00	.02	
Dynamic strength	58.01	59	14.95	1.00[a]	-.48
Static strength	55.03	59	17.93	1.00[a]	-.50
Both	55.21	58	17.75	1.00[a]	
Empirical two-dimensional					
Null	71.72	60	.00	.00	
General strength	53.19	59	18.53	1.00[a]	-.65
Knee strength	65.45	59	6.27	.44	-.46
Both	52.63	58	19.09	1.00[a]	
Empirical three-dimensional					
Null	68.06	60	.00	.00	
General strength	49.41	59	18.65	1.00[a]	-.65
Knee strength	62.40	59	5.66	.57	-.46
Handgrip strength	60.35	59	7.71	.83	-.49
All	48.54	57	19.52	1.00[a]	-.64
Seven-dimensional					
Null	68.13	60	.00	.00	
Upper body strength	54.17	59	13.96	1.00[a]	-.52
Trunk/leg extension strength	56.78	59	11.35	1.00[a]	-.46
Arm flexion strength	53.56	59	14.57	1.00[a]	-.52
Trunk flexion strength	56.63	59	11.50	1.00[a]	-.49
Handgrip strength	60.90	59	7.23	.76	-.35
Knee flexion strength	59.54	59	8.59	.93	-.40
Knee extension strength	57.67	59	10.46	1.00[a]	-.46
All seven	48.89	53	19.24	1.00[a]	
Upper body + arm flexion	53.54	58	14.59	1.00[a]	

[a]NNFI values >1.00 have been reported as 1.00.
[b]This is the correlation of the strength latent trait with the performance latent trait.

Strength and Performance

The analysis of the association of physical abilities and performance began by exploring the value of treating the strength traits defined by the various measurement models as causal influences on performance (Table 5). The χ^2 reduction relative to a model that assumed that strength and performance were independent was one criterion for choosing between models. A larger reduction indicated a better model. The correlation of the latent strength trait with the latent performance trait was a second criterion. A larger correlation indicated a better model. Model parsimony was a third criterion. A model with causal effect was preferable to a model with two causal effects.

Models based on the GS traits defined in the exploratory factor analyses were the best option (Table 5). The models that specified this latent trait as a cause of performance differences produced comparable χ^2 improvements whether the general dimension was defined by the two-dimensional model or the three-dimensional model. The χ^2 improvements approached the upper limit for any model in the table. Only models that involved multiple latent traits produced greater improvements in the fit of the model to the data. In each case, the parsimony principle supported the adoption of the GS model. The improvements in the model χ^2 were too small to justify adding causal effects to the model. The empirical GS models were also the best option by a third criterion. The correlation of the strength latent trait to performance was much stronger than in any other model ($r = .65$ vs. $r < .55$).

The evaluation of the seven-dimensional model deserves special comment. This model did not improve the overall ability to account for the associations of strength tests with performance. The last model fitted to the data included causal effects of each of the seven dimensions on performance. The χ^2 for this model ($\chi^2 = 19.24$) was only slightly larger than the χ^2 values for the models with the empirical GS dimensions as predictors ($\chi^2 = 18.52$ and $\chi^2 = 18.63$ for the two- and three-dimensional variants, respectively). The addition of 6 causal parameters to obtain a trivial χ^2 improvement was unreasonable.

The problem of choosing a model would not be any easier if the comparison had been limited to models specified within the seven-dimensional framework. Several models defined within this framework were roughly comparable. Five models produced χ^2 improvements between 10.46 and 14.57, with ability–performance correlations between $r = -.46$ and $r = .52$. The results for those five models were close enough to suggest that they could be considered equivalent models for practical purposes.

The GS dimensions from the empirical factor analyses provided the most reasonable representation of strength. The exclusion of knee strength measures is the primary difference between these latent traits and the a priori unidimensional model. That exclusion significantly improved the ability to predict performance.

Full Ability Model

The ability dimensions of ME, AP, and AC were added to the two- and three-dimensional empirical models. Analyses with the two models produce the same patterns of results, but the

three-dimensional model provided better overall predictive accuracy as would be expected from the analysis of strength dimensions in isolation from other abilities. Therefore, only the results for the three-dimensional model are considered here.

The models described in Table 6 can be compared using several criteria. How much does improve the fit of the model relative to the null model? The $\Delta\chi^2$ is the index of this criterion. How strong is the association of the ability trait with performance? The size of the latent trait correlations is the index for this criterion. How well does the model account for the overall association of abilities with performance? The number of modification indices for the excluded ability dimensions is the index for this criterion. For example, when considering the GS model,

Table 6. Performance Models for the Individual Dimensions in the Full Ability Model

	Model Evaluation Criteria				Modification Indices for Excluded Ability Dimensions					
	χ^2	$\Delta\chi^2$	*r*	*t*	GS	KS	HG	ME	AP	AC
Null	575.13									
GS	548.95	26.18	-.53	-4.38	--	.06	.69	.85	4.89	10.42
KS	563.62	11.51	-.40	-3.12	11.47	--	4.38	4.80	14.57	16.35
HG	558.09	17.04	-.43	-3.37	9.77	2.19	--	5.40	10.80	13.56
ME	563.60	11.53	-.46	-3.46	8.62	1.64	4.39	--	10.54	13.87
AP	543.80	31.33	-.54	-4.54	2.99	.91	.01	.67	--	2.72
AC	540.55	34.58	-.52	-4.51	7.69	1.84	1.72	3.12	2.09	--

Note. GS = general strength; KS = knee strength; HG = handgrip strength; ME = muscle endurance; AP = anaerobic power; AC = aerobic capacity.

only 2 of 5 modifications exceed 3.84, the minimum expected $\Delta\chi^2$ that would justify adding an ability factor to the existing model. The GS model would be superior to the KS model, which leaves all 5 modification indices above the critical value.

Table 6 also provides some insight into the necessity of including each ability in the final model. The index for this determination is the number of modification indices >3.84 in the column headed by the ability dimension. For example, GS would be considered for addition to 4 of 5 models, while there would be no reason to consider adding KS to any of the other models in the table.

The ability trait evaluation criteria split the ability traits into two groups. The first group consists of the models based on GS, AP, and AC. These models produced $\Delta\chi^2 > 25$, accounted for more than 25% of the performance variance ($r < -.51$), improved 4 of 5 models, and eliminated at least 4 of the other latent traits from consideration.

The second group consisted of models based on KS, HG, and ME. These models produced smaller improvements in fit ($\Delta\chi^2 < 20$), accounted for less than 22% of the performance variance ($r \geq -.46$), and produced substantial improvements in the fit of the model for at most 2

of the other 5 models. Finally, the models based on these dimensions eliminated no more than 2 of the other latent traits from consideration.

The evaluation of the single predictor models indicated that GS, AP, and AC are sufficiently similar to be considered competitive alternatives. The KS, HG, and ME models were not competitive with these three alternatives.

Closer examination of the GS, AP, and AC models indicated that none of them are consistently superior to the others. The $\Delta\chi^2$ criterion ranked the models AC>AP>GS. The strength of association criterion ranked the models AP>GS>AC. The model improvement criterion indicated GS=AP=AC. The trait elimination criterion ordered the models AP>AC>GS. Because there was no one dominant model, combinations of the GS, AP, and AC models were examined to define a final model (Table 7).

Table 7. Multivariate Models for Performance Prediction

	Model χ^2	$\Delta\chi^2$	r_i[a]	t_i[b]	r_2[a]	t_2[b]	R[c]	MI[d]
Null	575.13							
GS+AP	542.02	33.11	-.27	-1.53	-.34	-1.96	.57	7.33
GS+AC	534.43	40.70	-.33	-2.50	-.36	-2.81	.59	0.74
AP+AC	539.41	35.72	-.28	-1.16	-.30	-1.28	.55	7.05

[a]The *r*i is the correlation of the first (i = 1) or second (i = 2) latent trait listed in the model name.
[b]t_i is the *t* value for the first (i = 1) or second (i = 2) ability latent trait listed in the model name.
[c]Multiple correlation of the performance latent trait with the two ability latent traits.
[d]Modification index for the ability latent trait that was omitted from the model.

There were several reasons to prefer the GS+AC model when the three 2-predictor models were compared. This model produced the greatest improvement in fit relative to the null model. Both of the predictors were significantly related to the performance criterion; none of the associations were significant in the other models. The multiple correlation, *R*, was stronger than either of the other two models. The modification index for AP was trivial (MI = .74), whereas these indices would have justified adding a third predictor to either of the other two models (MI > 7.00).

Effect of Choice of AC Indicator

VO_2 in liters was the AC indicator in the initial model evaluations. Oxygen uptake scaled to body size, VO_{2max}, is often used as a predictor in performance studies. to predict in studies relating oxygen uptake to performance. The three models in Table 7 were compared a second time with VO_{2max} in ml/kg/min as the measure of AC (Table 8). In this comparison as in the previous one, the GS+AC model was the best choice by all criteria.

Table 8. Model Comparisons With VO_{2max} as the AC Indicator

								Modification Indices		
		$\Delta\chi^2$	r_1[a]	t_1[a]	r_2[a]	t_2[a]	R	GS	AP	AC
Null	571.71									
GS	544.84	26.97	-.53	-4.35			.53	--	3.21	15.07
AP	544.31	27.40	-.53	-4.22			.53	3.55	--	8.35
AC	551.95	19.76	-.44	-3.64			.44	23.22	16.18	--
GS+AP	541.24	30.47	-.31	-1.67	-.29	-1.53	.55	--	--	12.82
GS+AC	529.01	42.70	-.50	-4.50	-.42	-3.86	.63	--	.33	--
AP+AC	533.87	37.84	-.44	-3.65	-.33	-.291	.59	8.65	--	--

[a]The subscript indicates the correlation is for the first (1) or second (2) variable listed in the model name.

The GS+AC model with VO_{2max} as the AC indicator must be compared with the GS+AC with VO_2 in liters to determine the best available model. In this comparison, the VO_{2max} model performed better than the VO_2 in liters model by every comparison criterion. The standardized coefficients were larger, the associated *t* values were larger, the multiple correlation was larger, and the residual modification index for AP was smaller. The optimum model clearly was GS+AC with VO_{2max} as the AC indicator.

Residual Test–Task Associations

The residual associations of strength tests with task measures (4 tasks x 22 tests) were examined for the GS+AC (VO_{2max}) model. The examination was undertaken to determine how well the general model captured the overall association of tests with tasks. Four of the 88 standardized residuals were large enough to be statistically significant if just one residual had been examined (VO_{2max} – ammunition box lift, $z = -2.69$; trapezius endurance – jerry can carry, $z = -2.03$; VO_{2max} – jerry can carry, $z = -2.68$; leg peak power – slit trench digging, $z = 2.52$). To place this result in context, 39 of 88 standardized residuals were significant in the null model. Fourteen of the residuals for the null model were greater in absolute value than the maximum value ($z = 2.69$) seen in the final model. The maximum z-score for the residuals in the null model ($z = -4.35$) was much larger than the maximum in the final model. Further context is provided by considering that there is a better than 50:50 chance of finding 4 or more significant associations by chance when 88 tests are performed ($p = .52$). Finally, the Bonferroni adjustment to maintain the overall error rate at $p = .05$ for the set of 88 residuals is $p < .0006$, which corresponds to $z > 3.45$ for a 2-tailed test. None of the observed residuals approached this value.

Discussion

The strength measurement models produced three distinct sets of models. One set consisted of the unidimensional model, the isometric model, and the isokinetic model. A second set consisted of models based on five of the dimensions from the specific functions model. The third set consisted of the 9- and 11-test GS dimensions in the empirical models. These sets were defined by noting that the models within each set were about equally effective in accounting for

the covariation of test scores with task scores. The models within each set also were about equally effective in predicting performance.

The choice between the three sets of equivalent models was straightforward. The empirical GS dimensions models accounted for 42% of the performance variance (r = -.65). The models in the other two sets accounted for at most 27% of that variance (general dimensions, -.50 ≤ r ≤-.40; specific dimensions, -.52 ≤ r ≤-.35).

The initial steps in constructing the final performance prediction model identified GS, AP, and AC as the major correlates of performance. GS and AC were the only abilities in the final model. This pairing replicated Vickers et al.'s (2009) findings for moderate duration physical tasks (Vickers et al., 2009).

The study shed light on the belief that specific tests must be administered to predict task performance (McArdle, Katch, & Katch, 2001, p. 597). Following this logic, ME and AP would be expected to be the primary predictors of task performance. The rationale would be that the duration of the physical tasks in this study fell in a range that is generally thought of as requiring these abilities. However, the study by Vickers et al. (2009) also included ME and AP, and it too excluded those dimensions from the final model. These negative findings might be countered by arguing that test-task matching should be carried out at the level of individual tests and specific tasks. If this argument were correct, substantial test-task residuals would be expected. While some nominally significant (p < .05) residuals were found in this study, the number was no greater than expected by chance. MacCallum, Roznowski, & Necowitz (1992) pointed out that a substantial residual often is a chance event that does not replicate. Experience has shown that test-task residuals are not likely to replicate when modeling the relationships of physical abilities to physical tasks (Vickers, 1996; Vickers et al., 2009). Still, the residuals in this study may replicate in future work. Until then, there is no reason to invoke test-task specificity as a criterion for task prediction. One benefit of this conclusion is that the ability–performance model constructed here can be generalized tentatively to moderate duration tasks that were not covered in this study. Such generalization would not be possible with a test-task approach. Every task would require its own model. For the present, the important point is that subjective judgments based on test-task matching are likely to be a poor guide to identifying the key physical abilities for task performance.

The study extended an earlier demonstration of the risk of omitted variable bias in ability–performance modeling (Vickers et al., 2009). Bias arises when a causal model omits one or more causal factors and those omitted causes are correlated with predictors in the model. When this happens, the model will assign part of the causal effects of the omitted variables to variables that are in the model, thereby biasing the estimates for the included variables (James et al., 1982). The final model can be used to illustrate this risk. If HG, KE, ME, or AP had been studied in isolation, the correlation of each dimension with performance could have been interpreted as a causal effect. The final model implies that each of these apparent causal models would have been the spurious product of omitted variable bias. Any training program based on those spurious models would do little to improve performance. Because physical abilities often display moderate to strong correlations, omitted variable bias should always be considered when constructing causal models relating physical abilities to performance.

Several additional points merit brief comment. First, measurement modality had little effect on strength assessments. This result replicated an earlier finding by Dempsey et al. (1998). To date, isometric, isotonic or isoinertial, and isokinetic measures all are reasonable options for strength measurement. Second, a hierarchical strength measurement model might be appropriate. Vickers' (2003b) analysis of extensive batteries of tensiometer measures led to the same conclusion. The lowest level of the hierarchy would consist of specific strength factors. The intermediate level would distinguish upper strength from lower body strength. GS would be the highest level in the model. This hierarchy could combine the specific strength dimension model developed here with dimensions found in the empirical two-dimensional model and would yield a GS dimension based on the correlation between the upper and lower body strength dimensions. Third, the fact that size-adjusted maximal oxygen uptake, VO_{2max}, predicted performance better than absolute oxygen uptake, VO_{2L}, was surprising since VO_{2L} would seem to be the more natural indicator of how rapidly the energy demands of a fixed task could be met. The unanticipated finding may be related to recent observations that high VO_{2max} is associated with faster activation of aerobic processes during exercise (Kilding, Fysh, & Winter, 2007; Kilding, Winter, & Fish, 2006) and with a higher anaerobic threshold (Myers & Ashley, 1997; Paterson & Cunningham, 1999). Thus, VO_{2max} is an index of the ability to maintain a higher level of submaximal energy production and to reach that level more rapidly after beginning to work. The results obtained here may indicate that these factors are more important than the actual maximum aerobic energy that can be generated. Finally, a performance was adequately represented by a single latent trait. This result was unexpected since the individual performance tasks had been chosen to represent combat activities that were perceived to require different abilities. The unidimensional structure of performance suggests task duration may be the key to understanding which physical abilities are required. All of the tasks in this study were of moderate duration. Prior work indicates that brief maximal materials-handling tasks also define a single performance factor (Vickers, 1995, 1996), and that this factor requires a different combination of abilities than moderate duration tasks similar to those in this study (Vickers et al., 2009). These observations lead to the inference that systematic task sampling is a critical factor to consider when identifying abilities that define combat readiness.

In summary, the replication of an earlier GS+AC model for moderate duration tasks was the most important finding in this study. This replication included rejecting ME and AP as critical abilities for moderate duration tasks. The GS dimension in the present model was a broad construct defined by static and dynamic strength tests and encompassing both upper and lower body strength tests. The study reinforced doubts about the effectiveness of test-task specificity as a basis for causal inferences about the ability–performance interface. The study also reinforced concerns about omitted variable bias as a problem for performance modeling. The explanation of why GS+AC model works well remains uncertain, but the fact that this model no has proven to be the best option in each of two studies indicates that it is a reliable framework for identifying abilities to target in physical training programs undertaken to improve performance on moderate duration military tasks.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*, 411-423.

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago: SmallWaters Corporation.

Beckett, M. B., & Hodgdon, J. A. (1987). *Lifting and carrying capacities relative to physical fitness measures* (Technical Report No. 87-26). San Diego, CA: Naval Health Research Center.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.). *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Dempsey, P. G., Ayoub, M. M., & Westfall, P. H. (1998). Evaluation of the ability to predict low frequency lifting capacity. *Ergonomics,* 41, 1222-1241.

Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Erlbaum Associates.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179-185.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage Publications.

Joreskog, K. G., & Sorbom, D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kilding, A. E., Fysh, M., & Winter, E. M. (2007). Relationships between pulmonary oxygen uptake kinetics and other measures of aerobic fitness in middle- and long-distance runners. *European Journal of Applied Psychology, 100*(1), 105-114.

Kilding, A. E., Winter, E. M., & Fysh, M. (2006). Moderate-domain pulmonary oxygen uptake kinetics and endurance running performance. *Journal of Sports Sciences, 24*(9), 1013-1022.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490-504.

McArdle, W. D., Katch, F. I., & Katch, V. L. (2001). *Exercise physiology: Energy, nutrition, and human performance* (5th ed.). Philadelphia: Lippincott Williams and Wilkins.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*, 64-82.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1*(2), 108-141.

Myers, J., & Ashley, E. (1997). Dangerous curves. *Chest, 111*, 787-795.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*(3), 396-402.

Paterson, D. H., & Cunningham, D. A. (1999). The gas transporting systems: Limits and modifications with age and training. *Canadian Journal of Applied Physiology, 24*, 28-90.

Rayson, M. P., Holliman, D., & Belyavin, A. (2000). Development of physical selection procedures for the British Army. Phase 2: Relationship between physical performance tests and criterion tasks. *Ergonomics, 43*(1), 73-105.

Robertson, D. W., & Trent, T. (1982). *Validity of an occupational strength test battery (STB) for early identification of potential underwater demolition team and sea/air/land team trainees* (Technical Report No. NPRDC TR 84-2). San Diego, CA: Navy Personnel Research and Development Center.

Robertson, D. W., & Trent, T. T. (1985). *Documentation of muscularly demanding job tasks and validation of an occupational strength test battery (STB)* (Technical Report No. 86-1). San Diego, CA: Navy Personnel Research and Development Center.

Singh, M., Lee, S. W., Wheeler, G. D., Chahal, P., Oseen, M., & Couture, R. T. (1991). *Development of Forces Mobil Command Army physical fitness evaluation and standards for field units* (Final Report). Edmonton, Alberta, Canada: Faculty of Physical Education and Recreation, University of Alberta.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173-180.

Tucker, L. R. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38,1-10.

Vickers, R. R., Jr. (1995). *Physical task performance: Complexity of the ability–performance interface* (Technical Report No. 95-30). San Diego, CA: Naval Health Research Center.

Vickers, R. R., Jr. (1996). *Generalizability test of a physical ability–job performance model* (Technical Report No. 96-16). San Diego, CA: Naval Health Research Center.

Vickers, R. R., Jr. (2003a). *Physical strength and the performance of moderate duration tasks* (Technical Report No. 03-08). San Diego, CA: Naval Health Research Center.

Vickers, R. R., Jr. (2003b). *The measurement structure of muscle strength* (Technical Report No. 03-30). San Diego, CA: Naval Health Research Center.

Vickers, R. R., Jr., Hodgdon, J. A., & Beckett, M. A. (2009). *Physical ability–task performance models: Assessing the risk of omitted variable bias* (Technical Report No. 09-04. San Diego, CA: Naval Health Research Center.

Appendix. Descriptive Statistics From Singh et al. (1991)

Measure	*M*	*SD*
Anaerobic power indicators		
Leg peak power	762.8	113.0
Arm peak power	444.0	83.3
Aerobic capacity indicators		
VO_{2max} (L/min)	52.28	0.60
VO_{2max} (ml/kg/min)	52.28	6.31
AT VO_2 (ml/kg/min)	45.52	6.16
Isometric strength tests		
Right handgrip	55.3	8.2
Left handgrip	51.9	8.6
Arm flexion	46.7	15.8
Trunk flexion	62.9	10.7
Trunk extension	171.0	25.1
Isokinetic strength tests		
Right knee flexion	115.4	19.3
Left knee flexion	114.1	22.4
Right knee extension	157.7	26.6
Left knee extension	156.1	25.5
Arm flexion	77.2	20.4
Trunk flexion	73.0	10.6
Trunk extension	161.7	24.7
Leg extension	257.9	62.9
Trapezius lift	62.7	15.6
Bench press	116.4	26.4
Muscle endurance tests		
Static right handgrip (s)	119.9	52.3
Static left handgrip (s)	107.0	42.8
Static arm flexion (s)	109.3	43.9
Dynamic trapezius lift (reps)	92.5	20.1
Performance tasks		
Casualty evacuation (s)	46.82	8.52
Ammunition box lift	164.27	50.62
Jerry can carry	242.32	30.10
Digging slit trench	262.04	44.54

Note. Of 88 subjects who completed the weight-loaded march, 6 did not reach 50% of VO_{2max} and only 20 exceeded 70% of VO_{2max}. This test was definitely submaximal, so much so that the above information implies that most study participants did not reach their anaerobic threshold during the march.

REPORT DOCUMENTATION PAGE

1. Report Date (DD MM YY)
09 06 09

2. Report Type
Technical Report

3. DATES COVERED (from - to)
Apr 09 – Jun 09

4. TITLE AND SUBTITLE
Physical Abilities and Task Performance: A Replication and Extension

5a. Contract Number:
5b. Grant Number:
5c. Program Element:
5d. Project Number:
5e. Task Number:
5f. Work Unit Number: 60704

6. AUTHORS
Vickers, Ross. R., Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Commanding Officer
Naval Health Research Center
140 Sylvester Road
San Diego, CA 92106-3521

8. PERFORMING ORGANIZATION REPORT NUMBER
09-30

8. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

Commanding Officer
Naval Medical Research Center
503 Robert Grant Ave
Silver Spring, MD 20910-7500

Commander
Navy Medicine Support Command
P.O. Box 140
Jacksonville, FL 32213-0140

10. Sponsor/Monitor's Acronyms(s)
NMRC/NMSC

11. Sponsor/Monitor's Report Number(s)

12 DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT (maximum 200 words)

Strength influences the performance of military physical tasks. These influences can be summarized by models that treat strength as a general dimension that affects performance on tasks in general. Previous findings also indicate that combining a strength dimension with an aerobic capacity dimension yields a model that accounts for the full pattern of association between physical ability tests and lifting and carrying. This study attempted to replicate the earlier findings using a strength test battery with some new strength measures, a different set of military tasks, and a different military population. Structural equation models were constructed to represent strength as a single construct, a two-dimensional construct based on measurement modality, and a seven-dimensional model based on specific functional movements. Performance was represented as a single general performance dimension that added digging and casualty evacuation to the manual materials-handling tasks that had been previously studied. A modified unidimensional model maximized the prediction of performance. Adding aerobic capacity to the strength model improved performance prediction, but adding muscle endurance and anaerobic power did not. The results provided a very close replication of earlier findings while extending the model to a wider range of military tasks and a new population.

14. SUBJECT TERMS
physical ability, task performance, structural equation modeling

16. SECURITY CLASSIFICATION OF: a. REPORT	b.ABSTRACT	b. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UNCL	UNCL	UNCL	UNCL	21

18a. NAME OF RESPONSIBLE PERSON
Commanding Officer

18b. TELEPHONE NUMBER (INCLUDING AREA CODE)
COMM/DSN: (619) 553-8429

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18